

Claims

- [c1] 1. A method for extracting and structuring items of data from content available via the Internet, the method comprising:
- receiving input of a user specifying at least one source of content available via the Internet, types of data to be extracted from said at least one source, and fields for structuring extracted items of data;
- retrieving content from said at least one source;
- parsing the retrieved content to extract items of data of the types specified by the user; and
- mapping the extracted items of data to the fields specified by the user so as to transform the extracted items of data into a structured format.
- [c2] 2. The method of claim 1, wherein said at least one source of content includes a Web site.
- [c3] 3. The method of claim 1, wherein said at least one source of content includes an HTML page.
- [c4] 4. The method of claim 1, wherein said receiving step includes receiving a URL specifying a source of content available via the Internet.

- [c5] 5. The method of claim 1, wherein said receiving step includes receiving user input specifying attributes of data to be extracted from said at least one source.
- [c6] 6. The method of claim 1, wherein said retrieving step includes retrieving a Web page.
- [c7] 7. The method of claim 6, wherein said Web page comprises a selected one of an HTML page, an cHTML page, and an XHTML page.
- [c8] 8. The method of claim 6, wherein said parsing step includes parsing container objects of the Web page.
- [c9] 9. The method of claim 8, wherein said parsing step includes creating a new object for a particular container object of the Web page, the new object containing information for the particular container object.
- [c10] 10. The method of claim 8, wherein said parsing step includes creating feature tags for elements of the container objects.
- [c11] 11. The method of claim 10, wherein said creating feature tags step includes creating information for selected one of a headline, a graphic object, a button, and a run of text.

- [c12] 12. The method of claim 10, wherein said creating feature tags includes creating feature tags based on attributes of each element.
- [c13] 13. The method of claim 10, further comprising:
saving at least some of the feature tags; and
subsequently, using a saved feature tag to retrieve an element from the Web page.
- [c14] 14. The method of claim 1, wherein said mapping step includes receiving user input for selecting a particular field for placing a given item of data.
- [c15] 15. The method of claim 14, wherein said mapping step further comprises automatically mapping other items of data similar to the given item of data to the particular field.
- [c16] 16. The method of claim 1, wherein said mapping step includes generating an XML document including the extracted items of data.
- [c17] 17. The method of claim 1, wherein said mapping step includes saving the extracted information to a database table.
- [c18] 18. A computer-readable medium having processor-executable instructions for performing the method of

claim 1.

- [c19] 19. A downloadable set of processor-executable instructions for performing the method of claim 1.